CCIT Report #896 May 2016

On Empirical Cumulant Generating Functions of Code Lengths for Individual Sequences

Neri Merhav

Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, ISRAEL
E–mail: merhav@ee.technion.ac.il

Abstract

We consider the problem of lossless compression of individual sequences using finite–state (FS) machines, from the perspective of the best achievable empirical cumulant generating function (CGF) of the code length, i.e., the normalized logarithm of the empirical average of the exponentiated code length. Since the probabilistic CGF is minimized in terms of the Rényi entropy of the source, one of the motivations of this study is to derive an individual–sequence analogue of the Rényi entropy, in the same way that the FS compressibility is the individual–sequence counterpart of the Shannon entropy. We consider the CGF of the code-length both from the perspective of fixed–to–variable (F-V) length coding and the perspective of variable–to–variable (V-V) length coding, where the latter turns out to yield a better result, that coincides with the FS compressibility. We also extend our results to compression with side information, available at both the encoder and decoder. In this case, the V–V version no longer coincides with the FS compressibility, but results in a different complexity measure.

Index Terms Individual sequences, compressibility, finite–state machines, cumulant generating function, Rényi entropy, Lempel–Ziv algorithm.